

Partition bound is quadratically tight for product distributions*

Prahladh Harsha[†]

Rahul Jain[‡]

Jaikumar Radhakrishnan[§]

Abstract

Let $f : \{0, 1\}^n \times \{0, 1\}^n \rightarrow \{0, 1\}$ be a 2-party function. For every product distribution μ on $\{0, 1\}^n \times \{0, 1\}^n$, we show that

$$CC_{0.49}^\mu(f) = O\left(\left(\log \text{prt}_{1/8}(f) \cdot \log \log \text{prt}_{1/8}(f)\right)^2\right),$$

where $CC_\varepsilon^\mu(f)$ is the distributional communication complexity of f with error at most ε under the distribution μ and $\text{prt}_{1/8}(f)$ is the *partition bound* of f , as defined by Jain and Klauck [*Proc. 25th CCC*, 2010]. We also prove a similar bound in terms of $\text{IC}_{1/8}(f)$, the *information complexity* of f , namely,

$$CC_{0.49}^\mu(f) = O\left(\left(\text{IC}_{1/8}(f) \cdot \log \text{IC}_{1/8}(f)\right)^2\right).$$

The latter bound was recently and independently established by Kol [*Proc. 48th STOC*, 2016] using a different technique.

We show a similar result for query complexity under product distributions. Let $g : \{0, 1\}^n \rightarrow \{0, 1\}$ be a function. For every bit-wise product distribution μ on $\{0, 1\}^n$, we show that

$$QC_{0.49}^\mu(g) = O\left(\left(\log \text{qprt}_{1/8}(g) \cdot \log \log \text{qprt}_{1/8}(g)\right)^2\right),$$

where $QC_\varepsilon^\mu(g)$ is the distributional query complexity of f with error at most ε under the distribution μ and $\text{qprt}_{1/8}(g)$ is the *query partition bound* of the function g .

Partition bounds were introduced (in both communication complexity and query complexity models) to provide LP-based lower bounds for randomized communication complexity and randomized query complexity. Our results demonstrate that these lower bounds are polynomially tight for *product* distributions.

*Work done when the three authors were visiting the Simons Institute for the Theory of Computing, Berkeley, USA.

[†]TIFR, Mumbai. prahladh@tifr.res.in. Research supported in part by ISF-UGC grant 1399/4 and Google India Fellowship.

[‡]CQT, MajuLab and NUS, Singapore. rahul@comp.nus.edu.sg. Partly supported by the Singapore Ministry of Education via Academic Research Fund Tier 3 MOE2012-T3-1-009 and Young Researcher Award, National University of Singapore.

[§]TIFR, Mumbai. jaikumar@tifr.res.in.

1 Introduction

Over the last decade, several lower bound techniques using linear programming formulations and information complexity methods have been developed for problems in communication complexity and query complexity. One of the central questions in communication complexity is to understand the tightness of these lower bound techniques. For instance, over the last few years, considerable effort has gone into understanding the *information complexity* measure. Informally speaking, (internal) information complexity is the amount of information the two parties reveal to each other about their respective inputs while computing the joint function. It is known that for product distributions, the internal information complexity not only lower bounds but also upper bounds the distributional communication complexity (up to logarithmic multiplicative factors in the communication complexity) [2]. On the other hand, recent works due to Ganor, Kol and Raz [6, 7, 8] show that there exist non-product distributions which exhibit exponential separation between internal information complexity and distributional communication complexity¹. However, it is still open if internal information complexity (or a polynomial of it) upper bounds the public-coin randomized communication complexity (up to logarithmic multiplicative factors in the input size) [5].

Jain and Klauck [11], using tools from linear programming, gave a uniform treatment of several of the existing lower bound techniques and proposed the *partition bound*. This leads to following related (but incomparable) conjecture: does a polynomial of the partition bound yield an upper bound on the communication complexity? We are not aware of any counterexample to this conjecture².

We consider these questions when the inputs to Alice and Bob are drawn from a product distribution and show the following.

Theorem 1.1. *Let $f : \{0, 1\}^n \times \{0, 1\}^n \rightarrow \{0, 1\}$, and let $IC_\epsilon(f)$ and $\text{prt}_\epsilon(f)$ be the information complexity and partition bound respectively of f with error at most ϵ . For a product distribution μ on $\{0, 1\}^n \times \{0, 1\}^n$, the distributional communication complexity of f under distribution μ with error at most 0.49, denoted by $CC_{0.49}^\mu(f)$, can be bounded above as follows:*

$$CC_{0.49}^\mu(f) = O\left(\left(IC_{1/8}(f) \cdot \log IC_{1/8}(f)\right)^2\right), \quad (1.1)$$

$$CC_{0.49}^\mu(f) = O\left(\left(\log \text{prt}_{1/8}(f) \cdot \log \log \text{prt}_{1/8}(f)\right)^2\right). \quad (1.2)$$

Our technique yields bounds more general than those stated above (see discussion after [Proposition 2.5](#) for this generalization). We remark that recently (and independently of this work) Kol [14] obtained the bound (1.1) using very different techniques. Kol's result is stronger in the sense that her bound is in terms of the information complexity $IC^\mu(f)$ for the product distribution μ , while our result is in terms of the worst case information complexity $IC(f)$ (note, $IC_\epsilon(f) = \max_\mu IC_\epsilon^\mu(f)$). In fact, Kol showed that

$$CC_{\delta+\epsilon}^\mu(f) = O\left(IC_\delta^\mu(f)^2 \cdot \text{poly log } IC_\delta^\mu(f)/\epsilon^5\right), \quad (1.3)$$

¹The third result of Ganor, Kol and Raz [8] actually demonstrates an exponential separation between external information and communication complexity, albeit not for computing a Boolean function.

²The recent work of Göös et al. [9] demonstrates the existence of a total function for which the partition bound is strictly sublinear in the randomized communication complexity. This still does not rule out communication complexity being bound by a polynomial of the partition bound.

and concluded that

$$CC_{0.49}^\mu(f) = O\left(\text{IC}_{1/8}(f)^2 \cdot \text{poly log IC}_{1/8}(f)\right). \quad (1.4)$$

Kol's result (1.3) is incomparable to our second result in terms of partition bound (1.2).

We consider a similar question in query complexity and show the following.

Theorem 1.2. *Let $g : \{0,1\}^n \rightarrow \{0,1\}$ be a function and μ be a bit-wise product distribution on $\{0,1\}^n$. Let $\text{qprt}_\varepsilon(g)$ be the query partition bound for g with error ε . Then, the distributional query complexity with error at most 0.49 under the distribution μ , denoted by $QC_{0.49}^\mu(f)$, can be bounded above as follows:*

$$QC_{0.49}^\mu(g) = O\left(\left(\log \text{qprt}_{1/8}(g) \cdot \log \log \text{qprt}_{1/8}(g)\right)^2\right).$$

A similar quadratic upper bound for query complexity for product distributions in terms of approximate certificate complexity was obtained by Smyth [19]. His proof uses Reimer's inequality while our proof technique is based on Nisan and Wigderson's [17] more elementary approach.

Organization

The rest of the paper is devoted to the proofs of these two theorems. The communication complexity result is proven in Section 2 while the query complexity result is proved in Section 3.

2 Communication Complexity

2.1 Preliminaries

We work in Yao's two-party communication model [20] (see Kushilevitz and Nisan [15] for an excellent introduction to the area). Let \mathcal{X}, \mathcal{Y} and \mathcal{Z} be finite non-empty sets, and let $f : \mathcal{X} \times \mathcal{Y} \rightarrow \mathcal{Z}$ be a function. A two-party protocol for computing f consists of two parties, Alice and Bob, who get inputs $x \in \mathcal{X}$ and $y \in \mathcal{Y}$ respectively, and exchange messages in order to compute $f(x, y) \in \mathcal{Z}$ (using shared randomness).

For a distribution μ on $\mathcal{X} \times \mathcal{Y}$, let the ε -error distributional communication complexity of f under μ (denoted by $CC_\varepsilon^\mu(f)$), be the number of bits communicated (for the worst-case input) by the best deterministic protocol for f with average error at most ε under μ . Let $CC_\varepsilon^{\text{pub}}(f)$, the public-coin randomized communication complexity of f with worst case error ε , be the number of bits communicated (for the worst-case input) by the best public-coin randomized protocol that for each input (x, y) computes $f(x, y)$ correctly with probability at least $1 - \varepsilon$. Randomized and distributional complexity are related by the following special case of von Neumann's minmax principle.

Theorem 2.1 (Yao's minmax principle [21]). $CC_\varepsilon^{\text{pub}}(f) = \max_\mu CC_\varepsilon^\mu(f)$.

We will prove Theorem 1.1 by first showing an upper bound on communication complexity in terms of the smooth rectangle bound and then observing that the smooth rectangle bound is bounded above by the partition bound.

Smooth rectangle bound: The smooth rectangle bound was introduced by Jain and Klauck [11] as a generalization of the rectangle bound. Just like the rectangle bound, the smooth rectangle bound also provides a lower bound for randomized communication complexity. Informally, the smooth rectangle bound for a function f under a distribution μ , is the maximum over all functions g , which are close to f under the distribution μ , of the rectangle bound of g . However, it will be more convenient for us to work with the following linear programming formulation. (See [11, Lemma 2] and [12, Lemma 6] for the relations between the LP formulation and the more “natural” formulation in terms of rectangle bound.) It is evident from the LP formulation that the smooth rectangle bound is a further relaxation of the partition bound (defined in the appendix). We will formulate our results in terms of a distributional version of the above smooth rectangle bound. For $\mu : \mathcal{X} \times \mathcal{Y} \rightarrow \mathbb{R}$ and any $z \in \mathcal{Z}$ and rectangle R , let $\mu_z(R) := \mu(R \cap f^{-1}(z))$ and $\mu_{\bar{z}}(R) := \mu(R) - \mu_z(R)$. Furthermore, let $\mu_z := \mu_z(\mathcal{X} \times \mathcal{Y})$ and $\mu_{\bar{z}} := \mu_{\bar{z}}(\mathcal{X} \times \mathcal{Y})$. The smooth rectangle and its distributional version are defined below.

Definition 2.2 (Smooth rectangle bound).

- For a function $f : \mathcal{X} \times \mathcal{Y} \rightarrow \mathcal{Z}$ and $\varepsilon \in (0, 1)$, the (ε, δ) -smooth rectangle bound of f denoted $\text{srec}_{\varepsilon, \delta}(f)$ is defined to be $\max\{\text{srec}_{\varepsilon, \delta}^z(f) : z \in \mathcal{Z}\}$, where $\text{srec}_{\varepsilon, \delta}^z(f)$ is the optimal value of the following linear program.
- For a distribution μ on $\mathcal{X} \times \mathcal{Y}$ and function $f : \mathcal{X} \times \mathcal{Y} \rightarrow \mathcal{Z}$, the (ε, δ) -smooth rectangle bound of f with respect to μ denoted $\text{srec}_{\varepsilon, \delta}^\mu(f)$ is defined to be $\max\{\text{srec}_{\varepsilon, \delta}^{z, \mu}(f) : z \in \mathcal{Z}\}$, where $\text{srec}_{\varepsilon, \delta}^{z, \mu}(f)$ is the optimal value of the following linear program.

$$\begin{array}{ll}
 \text{srec}_{\varepsilon, \delta}^z(f) & \text{srec}_{\varepsilon, \delta}^{z, \mu}(f) \\
 \min \sum_R w_R & \min \sum_R w_R \\
 \sum_{R: (x, y) \in R} w_R \geq 1 - \varepsilon, & \sum_{(x, y) \in f^{-1}(z)} \mu_{x, y} \sum_{R: (x, y) \in R} w_R \geq (1 - \varepsilon) \cdot \mu_z \quad (2.1) \\
 \sum_{R: (x, y) \in R} w_R \leq \delta, & \sum_{R: (x, y) \in R} w_R \leq \delta, \quad \forall (x, y) \notin f^{-1}(z) \quad (2.2) \\
 \sum_{R: (x, y) \in R} w_R \leq 1, & \sum_{R: (x, y) \in R} w_R \leq 1, \quad \forall (x, y) \quad (2.3) \\
 w_R \geq 0, & w_R \geq 0, \quad \forall R.
 \end{array}$$

We will refer to the constraint in (2.1) as the covering constraint and the ones in (2.2) as the packing constraints. Note that while there is a single covering constraint (averaged over all the inputs (x, y) that satisfy $f(x, y) = z$) there are packing constraints corresponding to each $(x, y) \notin f^{-1}(z)$.

Similar to Yao’s minmax principle Theorem 2.1, we have the following proposition relating the distributional version of the smooth rectangle bound to the smooth rectangle bound.

Proposition 2.3. $\text{srec}_{\varepsilon, \delta}(f) = \max_{\mu} \text{srec}_{\varepsilon, \delta}^\mu(f)$.

The main result of this section is the following

Theorem 2.4. For any Boolean function $f : \{0, 1\}^n \times \{0, 1\}^n \rightarrow \{0, 1\}$ and any product distribution μ on $\{0, 1\}^n \times \{0, 1\}^n$, we have the following.

1.

$$\text{CC}_{0.49}^\mu(f) = O\left((\log \text{srec}_{1/n^2, 1/n^2}^\mu(f))^2 \cdot \log n\right).$$

2. Furthermore, if there exists $k \geq 20$ such that

$$\lceil 100 \log \text{srec}_{\delta, \delta}^\mu(f) \rceil \leq k,$$

for $\delta \leq 1/(30 \cdot 100(k+1)^4)$, then

$$\text{CC}_{0.49}^\mu(f) = O(k^2).$$

The above theorem is useful only when we have an upper bound on the smooth rectangle bound for very small δ . The following proposition shows that such upper bounds for smooth rectangle bound for such small δ can be obtained in terms of either the information complexity or the partition bound.

Proposition 2.5. For any Boolean function $f : \{0, 1\}^n \times \{0, 1\}^n \rightarrow \{0, 1\}$ and any $\delta \in (0, 1)$, we have the following bounds on $\text{srec}_{\delta, \delta}(f)$.

$$\log \text{srec}_{\delta, \delta}(f) \leq O\left(\log \frac{1}{\delta}\right) \cdot \text{IC}_{1/8}(f),$$

$$\log \text{srec}_{\delta, \delta}(f) \leq O\left(\log \frac{1}{\delta}\right) \cdot \log \text{prt}_{1/8}(f).$$

(This proposition depends on the error-reduction properties of information complexity and partition bound; a proof appears in [Appendix B](#).) Using this proposition, we can reduce the error (i.e., δ) to $1/n^2$ and show that $\text{CC}_{0.49}^\mu(f) = O\left((\log \text{prt}_{1/8}(f))^2 \cdot (\log n)^3\right)$. However, we can also reduce the error to $1/\text{poly}(\log \text{prt}_{1/8}(f))$ and show that there exists a $k = O(\log \text{prt}_{1/8}(f) \cdot \log \log \text{prt}_{1/8}(f))$ that satisfies the hypothesis for the second part of [Theorem 2.4](#). The bound (1.2) in [Theorem 1.1](#) now follows by combining [Propositions 2.3](#) and [2.5](#) and [Theorem 2.4](#). A similar argument yields the bound (1.1).

In particular, the above discussion shows that our techniques apply to any complexity measure (not necessarily partition bound and information complexity) which can be used to bound the smooth rectangle bound for very small δ . An interesting question that arises in this context is if we could bound smooth rectangle bound for small δ in terms of smooth rectangle bound for large δ , say $\delta = 1/3$ (i.e., is error-reduction for srec feasible?). This question was answered in the negative for partial functions by Göös et al. [10] who show that there exists a partial function f that has $\text{srec}_{1/3}(f) = O(\log n)$ and yet $\text{srec}_{1/4}(f) = \Omega(n)$.

2.2 Proof of [Theorem 2.4](#)

In this section, we construct a communication protocol tree with a small number of leaves from the optimal solutions to the LPs corresponding to $\text{srec}_{\varepsilon, \delta}^{0, \mu}$ and $\text{srec}_{\varepsilon, \delta}^{1, \mu}$. The construction of the protocol tree with a small number of leaves is inspired by a construction due to Nisan and Wigderson, in the context of log-rank conjecture [17, Theorem 2] (see also [15, Combinatorial proof of Theorem 2.11]).

Unlike the earlier constructions, our protocol works for a distribution and allows for error. As a result, the decomposition into sub-problems needs to be performed more carefully. This step critically uses the product nature of the distribution μ .

The decomposition is accomplished using an inductive argument. We will work with the quantity $\text{srec}^0 + \text{srec}^1$. That is, we will show that if this sum is small, then there is a protocol with few leaves. Suppose $\text{srec}^0 \leq \text{srec}^1$. Since srec^0 is small, we will conclude that there is a large rectangle biased towards 0 (see Lemma 2.6). Based on this large rectangle, the entire communication matrix is partitioned into three parts: (1) the large biased rectangle itself, (2) a rectangle whose corresponding sub-problem admits an LP solution leading to a smaller srec^1 value (the underlying product nature of the distribution μ is used here) and (3) a rectangle where the total measure with respect to μ drops significantly (see Lemma 2.7).

We say that a rectangle R is $(1 - \alpha)$ -biased towards 0 if $\mu_1(R) \leq \alpha\mu_0(R)$.

Lemma 2.6 (large biased rectangle). *Let μ be a product distribution. If $\text{srec}_{\varepsilon, \delta}^{0, \mu}(f) \leq D$, then for every $\rho \in (0, 1)$ there exists a rectangle S such that S is $(1 - \rho)$ -biased towards 0 and*

$$\mu(S) \geq \mu_0(S) \geq \frac{1}{D} \cdot \left((1 - \varepsilon) \cdot \mu_0 - \left(\frac{\delta}{\rho} \right) \cdot \mu_1 \right).$$

(The proof appears in Section 2.3.) We will apply the above lemma with $\rho = \sqrt{\delta}$ and conclude that there exists a large rectangle $S = X_0 \times Y_0$ that is $(1 - \sqrt{\delta})$ -biased towards 0. Let $X_1 = \mathcal{X} \setminus X_0$ and $Y_1 = \mathcal{Y} \setminus Y_0$. For $i, j \in \{0, 1\}$, define rectangles $R^{(ij)} := X_i \times Y_j$, $R^{(1*)} := X_1 \times \mathcal{Y}$, and $R^{(*1)} := \mathcal{X} \times Y_1$. (Note, $S = R^{(00)}$.) For $i, j \in \{0, 1, *\}$, let $\mu^{(ij)}$ be the restriction of μ to the rectangle $R^{(ij)}$. We show in the lemma below that the function f when restricted to either $R^{(10)}$ or $R^{(01)}$ has the property that the corresponding srec^1 drops by a constant factor. Define

$$\begin{aligned} \varepsilon(f) &:= 1 - \frac{\left(\sum_{(x,y) \in f^{-1}(1)} \mu_{x,y} \sum_{R:(x,y) \in R} w_R \right)}{\mu_1}, \\ \varepsilon^{(ij)}(f) &:= 1 - \frac{\left(\sum_{(x,y) \in f^{-1}(1) \cap R^{(ij)}} \mu_{x,y} \sum_{R:(x,y) \in R} w_R \right)}{\mu_1(R^{(ij)})}; \end{aligned} \quad \text{for } i, j \in \{0, 1\}.$$

It follows from the covering constraint that $\varepsilon(f) \leq \varepsilon$. Furthermore, $\varepsilon(f)$ is an average of the $\varepsilon^{(ij)}$'s in the sense that $\varepsilon(f) = \left(\sum_{i,j \in \{0,1\}} \mu_1(R^{(ij)}) \varepsilon^{(ij)} \right) / \mu_1$.

Lemma 2.7. *Suppose the product distribution μ and rectangles $R^{(ij)}$ are as above; in particular, $R^{(00)}$ is $(1 - \sqrt{\delta})$ -biased towards 0. There exists an $(ij) \in \{(01), (10)\}$ such that one of the following holds: (a) $2\mu^{(ij)}(f^{-1}(1)) \leq \mu^{(ij)}(f^{-1}(0))$ or (b) $\text{srec}_{\varepsilon^{(ij)} + 30\sqrt[4]{\delta}, \delta}^{1, \mu^{(ij)}}(f) \leq 0.9D$ where $\varepsilon^{(ij)}$ is as defined above.*

We will prove this lemma in Section 2.3. Let us assume the above lemmas and obtain the low cost communication protocol claimed in Theorem 2.4.

Suppose $\mu^{(01)}$ satisfies $\text{srec}_{\varepsilon^{(01)} + 30\sqrt[4]{\delta}, \delta}^{1, \mu^{(01)}}(f) \leq 0.9D$ as given by the above lemma. Consider the decomposition of the space $\mathcal{X} \times \mathcal{Y}$ given by $(R^{(00)}, R^{(01)}, R^{(1*)} = R^{(10)} \cup R^{(11)})$. We note that $R^{(00)}$ is a large biased rectangle, $R^{(01)}$ has lower srec^1 value while $R^{(1*)}$ has lower μ value (since $R^{(00)}$ is large) and its srec values are no larger than that of the entire space. In the case when $\mu^{(10)}$ satisfies $\text{srec}_{\varepsilon^{(10)} + 30\sqrt[4]{\delta}, \delta}^{1, \mu^{(10)}}(f) \leq 0.9D$, we similarly have the decomposition $(R^{(00)}, R^{(10)}, R^{(*1)} = R^{(01)} \cup R^{(11)})$.

This suggests a natural inductive protocol Π for f that we formalize in the lemma below.

For our induction it will be convenient to work with μ that are not necessarily normalized. So, we will only assume $\mu : \mathcal{X} \times \mathcal{Y} \rightarrow [0, 1]$ but not that $|\mu| := \mu(\mathcal{X} \times \mathcal{Y}) = \sum_{(x,y) \in \mathcal{X} \times \mathcal{Y}} \mu(x, y) = 1$. For a protocol Π , let the advantage of Π be defined by

$$\text{adv}_\mu(\Pi) = \sum_{(x,y): f(x,y)=\Pi(x,y)} \mu(x,y) - \sum_{(x,y): f(x,y) \neq \Pi(x,y)} \mu(x,y).$$

Let $L(\Pi)$ be the number of leaves in Π .

We now formulate the induction hypothesis as follows.

Lemma 2.8. *Fix a function $f : \mathcal{X} \times \mathcal{Y} \rightarrow \{0, 1\}$ and a product distribution (not necessarily normalized) $\mu : \mathcal{X} \times \mathcal{Y} \rightarrow [0, 1]$ such that $|\mu| \geq 0$. Let $\varepsilon, \delta \in (0, 1)$ and $\Delta \in (0, |\mu|)$. Let s, t be non-negative integers such that*

$$\begin{aligned} s &\geq s(\mu, \varepsilon, \delta) := \left\lceil 100 \cdot \log 2(\text{srec}_{\varepsilon, \delta}^{0, \mu}(f) + \text{srec}_{\varepsilon, \delta}^{1, \mu}(f)) \right\rceil; \\ t &\geq t(\mu, \varepsilon, \delta) := \lceil 100 \cdot 2^s \log(|\mu|/\Delta) \rceil. \end{aligned}$$

Then, there is a protocol Π such that

$$L(\Pi) \leq 4 \binom{s+t}{t} - 1; \tag{2.4}$$

$$\text{adv}_\mu(\Pi) \geq \left(\frac{1}{10} - \varepsilon - 30(s+1)^{4/3} \delta \right) |\mu| - \Delta \cdot L(\Pi). \tag{2.5}$$

Remark 2.9. Since $\varepsilon \leq \frac{1}{2}$, our definitions imply that $\text{srec}_{\varepsilon, \delta}^{1, \mu}(f) + \text{srec}_{\varepsilon, \delta}^{0, \mu}(f) \geq \frac{1}{2}$; thus $s \geq 0$. Similarly, since $\Delta \leq |\mu|$, we have $t \geq 0$.

Proof. First, we observe that if $\max\{\mu_0, \mu_1\} \geq 2 \min\{\mu_0, \mu_1\}$, then the protocol Π consisting of just one leaf, with the most popular value as label, meets the requirements: for, $\text{adv}_\mu(\Pi) \geq \frac{1}{3}|\mu|$ and $L(\Pi) = 1$, and our claim holds. Also, we may assume that $\varepsilon - 30(s+1)^{4/3} \delta < \frac{1}{10}$, for otherwise the claim is trivially true.

We now proceed by induction on $s+t$, assuming that μ is balanced: $\max\{\mu_0, \mu_1\} \leq 2 \min\{\mu_0, \mu_1\}$.

Base case ($s = 0$): Since $s = 0$, we have $\log \text{srec}_{\varepsilon, \delta}^{1, \mu}(f) \leq \frac{1}{100}$. We will show a protocol Π where Alice sends one bit after which Bob announces the answer. Consider the optimal solution $\langle w_R : R \text{ a rectangle} \rangle$ to the LP corresponding to $\text{srec}_{\varepsilon, \delta}^{1, \mu}(f)$; thus, $\text{OPT} := \sum_R w_R = \text{srec}_{\varepsilon, \delta}^{1, \mu}(f) \leq 2^{1/100} \leq 2$. Let $R = R_X \times R_Y$ be a random rectangle picked with probability proportional to w_R (using public coins). In the protocol Π , Alice tells Bob if $x \in R_X$, and Bob returns the answer 1 if $(x, y) \in R_Y$ and returns 0 otherwise. Let $p_{xy} := \Pr_R[(x, y) \in R]$. Then, by (2.1) we have $\sum_{(x,y) \in f^{-1}(1)} \mu(x, y) p_{xy} \geq (1 - \varepsilon)\mu_1/\text{OPT}$, and by (2.2), we have $\sum_{(x,y) \in f^{-1}(0)} \mu(x, y) p_{xy} \leq \delta\mu_0/\text{OPT}$. Thus,

$$\begin{aligned} \mathbb{E}_R \left[\sum_{(x,y): \Pi(x,y) \neq f(x,y)} \mu(x,y) \right] &= \sum_{(x,y) \in f^{-1}(1)} \mu(x,y)(1 - p_{xy}) + \sum_{(x,y) \in f^{-1}(0)} \mu(x,y)p_{xy} \\ &\leq \mu_1 - (1 - \varepsilon)\mu_1/\text{OPT} + \delta\mu_0/\text{OPT} \\ &\leq \mu_1 - ((1 - \varepsilon)\mu_1 - \delta\mu_0)/\text{OPT} \\ &\leq \frac{1}{2}(\mu_1 + \varepsilon\mu_1 + \delta\mu_0) \quad (\text{since } \text{OPT} \leq 2). \end{aligned} \tag{2.6}$$

Fix a choice R for which the quantity under the expectation is at most $\frac{1}{2}(\mu_1 + \varepsilon\mu_1 + \delta\mu_0)$. Then,

$$\begin{aligned} \text{adv}(\Pi) &= |\mu| - 2 \sum_{(x,y): \Pi(x,y) \neq f(x,y)} \mu(x,y) \\ &\geq |\mu| - (\mu_1 + \varepsilon\mu_1 + \delta\mu_0) \\ &\geq \left(\frac{1}{3} - \varepsilon - \delta\right) |\mu| \quad (\text{since } \mu_1 \leq 2\mu_0). \end{aligned}$$

Base case ($t = 0$): In this case, $|\mu| = \Delta$, and the protocol Π with a single leaf that gives the most probable answer achieves $\text{adv}(\Pi) \geq 0 \geq |\mu| - \Delta$.

Induction step: We will use [Lemma 2.6](#) to decompose the communication matrix into a small number of rectangles. After an exchange of a few bits to determine in which rectangle the input lies, Alice and Bob will be left with a problem for which s or t is significantly smaller. Assume $\text{srec}_{\varepsilon,\delta}^{1,\mu}(f) \geq \text{srec}_{\varepsilon,\delta}^{1,\mu}(f)$; in particular, $\text{srec}_{\varepsilon,\delta}^{1,\mu}(f) \leq 2^{s/100}$.

Formally, from [Lemma 2.6](#) (taking $\rho = \sqrt{\delta}$), we obtain a rectangle $R^{(00)} = X_0 \times Y_0$ such that (a) $R^{(00)}$ is $(1 - \sqrt{\delta})$ -biased towards 0, and (b) $\mu(R^{(00)}) \geq \frac{1}{2^{s/100}}(1 - \varepsilon - 2\sqrt{\delta})|\mu_0| \geq \frac{1}{3 \cdot 2^{s/100}}(1 - \varepsilon - 2\sqrt{\delta})|\mu|$. Recall the definitions of the rectangles $R^{(10)}, R^{(01)}, R^{(11)}, R^{(1*)}, R^{(*1)}$ and the corresponding restrictions of μ , namely, $\mu^{(01)}, \mu^{(10)}, \mu^{(11)}, \mu^{(1*)}, \mu^{(*1)}$. Suppose the choice of ij in [Lemma 2.7](#) for which one of the alternatives holds is $ij = 01$ (the other case $ij = 10$ is symmetric). The protocol Π proceeds as follows. Alice starts by telling Bob if $x \in X_0$.

Alice says $x \in X_0$. Now, Bob tells Alice if $y \in Y_0$.

Bob says $y \in Y_0$. The protocol $\Pi^{(00)}$ in this case has one leaf with answer 0; thus $\text{adv}(\Pi^{(00)}) \geq |\mu^{(00)}| \cdot (1 - \sqrt{\delta})$.

Bob says $y \notin Y_0$. Alice and Bob follow the protocol $\Pi^{(01)}$ promised by induction for $R^{(01)}$ under $\mu^{(01)}$. To bound the number of leaves in $\Pi^{(01)}$, we will consider the two alternatives ((a) and (b)) specified in [Lemma 2.7](#) separately. First (alternative (a)) suppose $2\mu^{(01)}(f^{-1}(1)) \leq \mu^{(01)}(f^{-1}(0))$; then we immediately declare 0 as the response, so that $L(\Pi^{(01)}) = 1$ and $\text{adv}(\Pi^{(01)}) \geq |\mu^{(01)}|/3$. If alternative (b) holds, then we have

$$\text{srec}_{\varepsilon^{(01)} + 30\sqrt[4]{\delta}, \delta}^{1,\mu^{(01)}}(f) \leq 0.9 \text{srec}_{\varepsilon,\delta}^{1,\mu}(f). \quad (2.7)$$

Then, we obtain $\Pi^{(01)}$ by induction. We take $\varepsilon^{(01)} + 30\sqrt[4]{\delta}$ as ε (if this quantity is greater than 1, then we use a trivial protocol with one leaf and zero advantage). With the reduction promised in (2.7), we may use a value of s that is the old s minus 1. Thus, we have

$$\begin{aligned} L(\Pi^{(01)}) &\leq 4 \binom{(s-1)+t}{t} - 1; \\ \text{adv}(\Pi^{(01)}) &\geq |\mu^{(01)}| \cdot \left(\frac{1}{10} - (\varepsilon^{(01)} + 30\sqrt[4]{\delta}) - 30s\sqrt[4]{\delta} \right) - \Delta \cdot L(\Pi^{(01)}). \end{aligned}$$

Alice says $x \notin X_0$. Alice and Bob follow the protocol $\Pi^{(1*)}$ obtained by applying the induction hypothesis to the rectangle $R^{(1*)}$ and the associated distribution $\mu^{(1*)}$. Observe that

$$|\mu^{(1*)}| \leq |\mu| - \mu(R^{(00)}) \leq |\mu| \left(1 - \frac{1}{3 \cdot 2^{s/100}}(1 - \varepsilon - 2\sqrt{\delta})\right) \leq |\mu| \left(1 - \frac{1}{4 \cdot 2^s}\right). \quad (2.8)$$

For the last inequality we used $\varepsilon + 2\sqrt{\delta} \leq \frac{1}{10}$, for otherwise [Eq. \(2.5\)](#) holds trivially. Now, [Eq. \(2.8\)](#) implies that $\log |\mu^{(1*)}| \leq \log |\mu| - \frac{1}{1002^s}$; so, for our induction we may take $t \leftarrow t - 1$. The parameters ε , δ and Δ remain the same. The original LP solutions are still valid for the subproblem, so we use the same s . The protocol $\Pi^{(1*)}$ obtained by induction satisfies the following inequalities.

$$\begin{aligned} L(\Pi^{(1*)}) &\leq 4 \binom{s + (t-1)}{t-1} - 1; \\ \text{adv}(\Pi^{(1*)}) &\geq |\mu^{(1*)}| \cdot \left(\frac{1}{10} - \varepsilon^{(1*)} - 30(s+1)\sqrt[4]{\delta}\right) - \Delta \cdot L(\Pi^{(1*)}). \end{aligned}$$

Putting all the contributions together, we obtain

$$\begin{aligned} L(\Pi) &= 1 + L(\Pi^{(01)}) + L(\Pi^{(1*)}) \\ &\leq 1 + \left(4 \binom{(s-1) + t}{t} - 1\right) + \left(4 \binom{s + (t-1)}{t-1} - 1\right) \\ &= 4 \binom{s + t}{t} - 1; \\ \text{adv}(\Pi) &\geq |\mu^{(00)}| \cdot (1 - \sqrt{\delta}) \\ &\quad + |\mu^{(01)}| \cdot \left(\frac{1}{10} - (\varepsilon^{(01)} + 30\sqrt[4]{\delta}) - 30s\sqrt[4]{\delta}\right) - \Delta \cdot L(\Pi^{(01)}) \\ &\quad + |\mu^{(1*)}| \cdot \left(\frac{1}{10} - \varepsilon^{(1*)} - 30(s+1)\sqrt[4]{\delta}\right) - \Delta \cdot L(\Pi^{(1*)}) \\ &\geq \left(\frac{1}{10} - \varepsilon - 30(s+1)\sqrt[4]{\delta}\right) |\mu| - \Delta \cdot L(\Pi). \quad \square \end{aligned}$$

The above lemma yields a protocol whose protocol tree has a small number of leaves, but not necessarily small depth. We can balance the protocol tree using the following proposition.

Proposition 2.10 ([15, Lemma 2.8]). *If f has a deterministic communication protocol tree with ℓ leaves, then f has a protocol tree with depth at most $O(\log \ell)$.*

We are now in a position to complete the proof of the main theorem of this section.

Proof of Theorem 2.4. To prove the first part of [Theorem 2.4](#), we invoke [Lemma 2.8](#) with $\Delta = 1/2^{4n}$ and $\varepsilon = \delta = 1/n^2$ to derive a protocol tree Π with at most

$$L(\Pi) = n^{O\left(\log \text{sec}_{1/n^2, 1/n^2}^{1, \mu}(f)\right)^2}$$

leaves and advantage at least $1/20$. The first part now follows from [Proposition 2.10](#).

To prove the second part of [Theorem 2.4](#), we invoke [Lemma 2.8](#) with $s = k$, $\Delta = 1/2^{5k^2}$ and $\varepsilon = \delta = 1/(30 \cdot 100(k+1)^4)$ where k satisfies the hypothesis. With this setting of parameters $t = \lceil 500 \cdot 2^k k^2 \rceil \leq 2^{2k}$ (for $k \geq 20$). [Lemma 2.8](#) implies a protocol tree Π with at most

$$L(\Pi) \leq (t + s)^s \leq t^{2s} \leq 2^{4k^2}$$

leaves and advantage at most $1/20$. The second claim then follows from [Proposition 2.10](#). \square

2.3 Proofs of [Lemmas 2.6](#) and [2.7](#)

Proof of [Lemma 2.6](#). Fix $z \in \{0, 1\}$. In the following we say that a rectangle R is *biased* (towards 0) if $\mu_1(R) \leq \rho \cdot \mu_0(R)$; otherwise, we say it is unbiased. Fix a solution $\langle w_R : R \text{ is a rectangle} \rangle$ that achieves the optimum $\text{srec}_{\varepsilon, \delta}^{0, \mu}(f) \leq D$. It follows

$$\begin{aligned} \sum_{R: \text{unbiased}} w_R \cdot \mu_0(R) &\leq \sum_{R: \text{unbiased}} w_R \cdot \frac{\mu_1(R)}{\rho} \\ &\leq \frac{1}{\rho} \cdot \sum_R w_R \cdot \mu_1(R) \\ &= \frac{1}{\rho} \sum_{(x, y) \in f^{-1}(1)} \mu(x, y) \sum_{R: (x, y) \in R} w_R \\ &\leq \frac{\delta}{\rho} \cdot \mu_1, \end{aligned}$$

where the last inequality follows from the packing constraints [\(2.2\)](#). We now use the covering constraints [\(2.1\)](#) to conclude that

$$\sum_{R: \text{biased}} w_R \cdot \mu_0(R) = \sum_R w_R \cdot \mu_0(R) - \sum_{R: \text{unbiased}} w_R \cdot \mu_0(R) \geq (1 - \varepsilon) \cdot \mu_0 - \frac{\delta}{\rho} \cdot \mu_1. \quad (2.9)$$

Define a probability distribution on the rectangles R as follows $p(R) := w_R / \text{srec}_{\varepsilon, \delta}^{0, \mu}(f)$. Then [\(2.9\)](#) can be rewritten as

$$\mathbb{E}_R [\mathbb{1}_{\text{biased}}(R) \cdot \mu_0(R)] \geq \frac{1}{D} \cdot \left((1 - \varepsilon) \cdot \mu_0 - \frac{\delta}{\rho} \cdot \mu_1 \right).$$

Hence, there exists a large biased rectangle $S = X_0 \times Y_0$ as claimed. \square

Proof of [Lemma 2.7](#). Since $R^{(00)}$ is $(1 - \sqrt{\delta})$ -biased towards 0, we have from the packing and covering constraints [\(2.2\)](#) and [\(2.3\)](#) that

$$\begin{aligned} \sum_{(x, y) \in R^{(00)}} \mu_{x, y} \sum_{R \ni (x, y)} w_R &= \sum_{(x, y) \in R^{(00)} \cap f^{-1}(1)} \mu_{x, y} \sum_{R \ni (x, y)} w_R + \sum_{(x, y) \in R^{(00)} \cap f^{-1}(0)} \mu_{x, y} \sum_{R \ni (x, y)} w_R \\ &\leq \mu_1(R^{(00)}) + \delta \mu_0(R^{(00)}) \leq (\sqrt{\delta} + \delta) \mu_0(R^{(00)}) \leq 2\sqrt{\delta} \mu(R^{(00)}). \end{aligned}$$

Hence,

$$\sum_R w_R \left(\frac{\mu(R^{(00)} \cap R)}{\mu(R^{(00)})} \right) \leq 2\sqrt{\delta}. \quad (2.10)$$

Group the rectangles in to subsets as follows:

$$B^{(01)} := \left\{ R : \frac{\mu(R^{(01)} \cap R)}{\mu(R^{(01)})} \geq \frac{10\sqrt[4]{\delta}}{D} \right\}, \quad B^{(10)} := \left\{ R : \frac{\mu(R^{(10)} \cap R)}{\mu(R^{(10)})} \geq \frac{10\sqrt[4]{\delta}}{D} \right\},$$

$$B := \left\{ R : \frac{\mu(R^{(11)} \cap R)}{\mu(R^{(11)})} \geq \frac{10}{D} \right\}.$$

By (2.3), we have

$$\sum_{(x,y) \in R^{(11)}} \mu_{x,y} \sum_{R \ni (x,y)} w_R \leq \sum_{(x,y) \in R^{(11)}} \mu_{x,y} = \mu(R^{(11)}).$$

Or equivalently,

$$\sum_R \frac{w_R}{D} \cdot \frac{\mu(R^{(11)} \cap R)}{\mu(R^{(11)})} \leq \frac{1}{D}.$$

Hence,

$$\sum_{R \in B} w_R \leq 0.1D. \quad (2.11)$$

We will now argue that either $\sum_{R \in B^{(01)}} w_R \leq 0.9D$ or $\sum_{R \in B^{(10)}} w_R \leq 0.9D$. Suppose, for contradiction, that neither is true. Then, by (2.11) we have

$$\sum_{R \in B^{(01)} \cap B^{(10)} \cap \bar{B}} w_R \geq 0.7D. \quad (2.12)$$

Since μ is a product distribution we have

$$\frac{\mu(R^{(01)} \cap R)}{\mu(R^{(01)})} \cdot \frac{\mu(R^{(10)} \cap R)}{\mu(R^{(10)})} = \frac{\mu(R^{(00)} \cap R)}{\mu(R^{(00)})} \cdot \frac{\mu(R^{(11)} \cap R)}{\mu(R^{(11)})}.$$

Using the above we have

$$\begin{aligned} \sum_R w_R \left(\frac{\mu(R^{(00)} \cap R)}{\mu(R^{(00)})} \right) &\geq \sum_{R \in B^{(01)} \cap B^{(10)} \cap \bar{B}} w_R \left(\frac{\mu(R^{(00)} \cap R)}{\mu(R^{(00)})} \right) \\ &\geq \sum_{R \in B^{(01)} \cap B^{(10)} \cap \bar{B}} w_R \left(\frac{\mu(R^{(01)} \cap R)}{\mu(R^{(01)})} \right) \cdot \left(\frac{\mu(R^{(10)} \cap R)}{\mu(R^{(10)})} \right) \bigg/ \left(\frac{\mu(R^{(11)} \cap R)}{\mu(R^{(11)})} \right) \\ &\geq \sum_{R \in B^{(01)} \cap B^{(10)} \cap \bar{B}} w_R \left(\frac{10\sqrt[4]{\delta}}{D} \right) \cdot \left(\frac{10\sqrt[4]{\delta}}{D} \right) \bigg/ \left(\frac{10}{D} \right) \\ &\geq \frac{10\sqrt{\delta}}{D} \cdot (0.7D) \\ &= 7\sqrt{\delta}. \end{aligned}$$

This contradicts (2.10). Hence, either $\sum_{R \in B^{(01)}} w_R \leq 0.9D$ or $\sum_{R \in B^{(10)}} w_R \leq 0.9D$. Assume, wlog that $\sum_{R \in B^{(01)}} w_R \leq 0.9D$. If f is 1/2-biased towards 0 with respect to the distribution $\mu^{(01)}$, then

the alternative (a) of the lemma holds, and we are done. Otherwise, that is $\mu_0(R^{(01)}) \leq 2\mu_1(R^{(01)})$ or equivalently $\mu(R^{(01)}) \leq 3\mu_1^{(01)}(R^{(01)})$. We will infer from this that $\text{srec}_{\varepsilon^{(01)} + 30\sqrt[4]{\delta}, \delta}^{1, \mu^{(01)}}(f) \leq 0.9D$. Consider the primal solution given by

$$w'_R = \begin{cases} w_R, & \text{if } R \in B^{(01)} \\ 0, & \text{if } R \notin B^{(01)}. \end{cases}$$

Clearly, w'_R , being a part of the original solution, satisfies (2.2) and (2.3), and has objective value at most $0.9D$. All we need to show is that it satisfies the covering constraint (2.1). For this, we first consider

$$\sum_{R \in B^{(01)}} w_R \left(\frac{\mu_1(R^{(01)} \cap R)}{\mu(R^{(01)})} \right) \leq \sum_{R \in B^{(01)}} w_R \left(\frac{\mu(R^{(01)} \cap R)}{\mu(R^{(01)})} \right) \leq \frac{10\sqrt[4]{\delta}}{D} \cdot D \leq 10\sqrt[4]{\delta}. \quad (2.13)$$

Now,

$$\begin{aligned} & \sum_{(x,y) \in f^{-1}(1) \cap R^{(01)}} \mu_{x,y} \sum_{R \in (x,y)} w'_R \\ &= \sum_{(x,y) \in f^{-1}(1) \cap R^{(01)}} \mu_{x,y} \sum_{R \in (x,y), R \in B^{(01)}} w_R \\ &= \sum_{(x,y) \in f^{-1}(1) \cap R^{(01)}} \mu_{x,y} \left(\sum_{R \in (x,y)} w_R - \sum_{R \in (x,y), R \notin B^{(01)}} w_R \right) \\ &= (1 - \varepsilon^{(01)})\mu_1(R^{(01)}) - \sum_{(x,y) \in f^{-1}(1) \cap R^{(01)}} \mu_{x,y} \sum_{R \in (x,y), R \notin B^{(01)}} w_R \\ &= (1 - \varepsilon^{(01)})\mu_1(R^{(01)}) - \sum_{R \notin B^{(01)}} w_R \mu_1(R^{(01)} \cap R) \\ &\geq (1 - \varepsilon^{(01)})\mu_1(R^{(01)}) - 10\sqrt[4]{\delta}\mu(R^{(01)}) && \text{[From (2.13)]} \\ &\geq (1 - \varepsilon^{(01)})\mu_1(R^{(01)}) - 30\sqrt[4]{\delta}\mu_1(R^{(01)}) && \text{[Since } \mu(R^{(01)}) \leq 3\mu_1(R^{(01)})\text{]} \\ &= (1 - \varepsilon^{(01)} - 30\sqrt[4]{\delta})\mu_1(R^{(01)}) \end{aligned}$$

Thus, (2.1) holds for $R^{(01)}$ with ε replaced by $\varepsilon^{(01)} + 30\sqrt[4]{\delta}$. \square

3 Query Complexity

Let $f : \{0, 1\}^n \rightarrow \{0, 1\}$ be the function for which we wish to build a decision tree.

3.1 Preliminaries

Definition 3.1 (product distribution). *We say $\mu : \{0, 1\}^n \rightarrow [0, 1]$ is a (bit-wise) product distribution on $\{0, 1\}^n$ if for $i = 1, 2, \dots, n$, there exist $p_i(0), p_i(1) \in [0, 1]$ (satisfying $p_i(0) + p_i(1) = 1$) such that for all $x \in \{0, 1\}^n$, $\mu(x) = \prod_i p_i(x_i)$.*

Let μ be a bit-wise product distribution on the inputs to f . Our goal is to build an efficient decision tree \mathcal{T} for f such that $\Pr_\mu[f(x) \neq \mathcal{T}(x)]$ is small. As the input bits are queried, the input is restricted to reside in a subcube. Let $s \in \{0, 1, \star\}^n$. The subcube of $\{0, 1\}^n$ with support s is

$$\text{subcube}(s) := \{x \in \{0, 1\}^n : s_i \neq \star \Rightarrow x_i = s_i\}.$$

Its size is $\text{size}(s) := |\{i : s_i \in \{0, 1\}\}|$. If A is a subcube, say $A = \text{subcube}(s)$, then $|A| := \text{size}(s)$.

We will derive an efficient decision tree in terms of the *query partition bound* of f due to Jain and Klauck [11]. The relationship of the query partition bound with other LP based bounds is described in [Appendix A](#).

Definition 3.2 (query partition bound [11]). Let $\varepsilon > 0$. In the following, A represents a subcube of $\{0, 1\}^n$. The ε -query partition bound of $f : \{0, 1\}^n \rightarrow \{0, 1\}$, denoted $\text{qpbt}_\varepsilon(f)$, is the optimal value of the following linear program.

$$\begin{aligned} \min \quad & \sum_z \sum_A w_{z,A} \cdot 2^{|A|} \\ & \sum_{A:x \in A} w_{f(x),A} \geq 1 - \varepsilon, & \forall x \in \{0, 1\}^n \\ & \sum_{A:x \in A} \sum_z w_{z,A} = 1, & \forall x \in \{0, 1\}^n \\ & w_{z,A} \geq 0, & \forall (z, A) \end{aligned}$$

Using the following claim (see appendix) one can ensure that the error parameter ε above is small with a modest increase in the query partition bound.

Claim 3.3. Let $\varepsilon > \delta > 0$. Then $\log \text{qpbt}_\delta(f) \leq \left(\frac{2}{(0.5-\varepsilon)^2} \log \frac{1}{\delta} \right) \log \text{qpbt}_\varepsilon(f)$.

Definition 3.4. We say μ is an $(\alpha_0, \beta_0, \alpha_1, \beta_1, a, b)$ -feasible distribution for f if there exists a feasible solution to the following inequalities. The variables are

$$(u_R : R \text{ is a subcube with support of size at most } a)$$

and

$$(w_R : R \text{ is a subcube with support of size at most } b).$$

$$\sum_{R:x \in R} u_R \geq 1 - \alpha_0 \quad \forall x \in f^{-1}(0) \quad (3.1)$$

$$\sum_{R:x \in R} u_R \leq \beta_0 \quad \forall x \in f^{-1}(1); \quad (3.2)$$

$$u_R \geq 0. \quad (3.3)$$

$$\sum_R \mu_1(R) w_R \geq (1 - \alpha_1) \mu_1 \quad (3.4)$$

$$\sum_{R:x \in R} w_R \leq 1 \quad \forall x \in \{0, 1\}^n; \quad (3.5)$$

$$\sum_{R:x \in R} w_R \leq \beta_1 \quad \forall x \in f^{-1}(0); \quad (3.6)$$

$$w_R \geq 0. \quad (3.7)$$

Remark 3.5. If the i -th bit of the input is fixed in μ (that is, $p_i(0) \in \{0, 1\}$), then we will assume i is not part of the support of any R in the above feasible solution.

The main technical contribution of this section is the following.

Lemma 3.6. *Let $\delta > 0$ and let μ be a product distribution that is $(\alpha_0, \beta_0, \alpha_1, \beta_1, a, b)$ -feasible for f . Then, there is a decision tree for f of depth at most ab that errs with probability at most*

$$\frac{1}{4} + \alpha_1 + \beta_1 + 4b(\beta_1 + \delta) + \frac{\beta_0}{(1 - \alpha_0)\delta}.$$

The query complexity bound stated in the introduction ([Theorem 1.2](#)) follows from the above lemma.

Proof of Theorem 1.2. Let $c := 8 + \log \text{qprt}_\varepsilon(f)$. Let $\gamma = 1/c^8$. Then from [Claim 3.3](#) we get that $d := \log \text{qprt}_\gamma(f) = O(c \log c)$. Let $\{w_{z,A}\}$ be an optimal solution for the primal of $\text{qprt}_\delta(f)$. Let $\mathcal{B} := \{A : |A| > d + \log \frac{1}{\gamma}\}$. Then $\sum_z \sum_{A \in \mathcal{B}} w_{z,A} < \gamma$ since $\sum_z \sum_A w_{z,A} \cdot 2^{|A|} = 2^d$. This implies (by first boosting and then removing the $A \in \mathcal{B}$) that μ is an $(\alpha_0, \beta_0, \alpha_1, \beta_1, a, b)$ -feasible distribution for f , with $\alpha_0 = \beta_0 = \alpha_1 = \beta_1 = 2\gamma$ and $a = b = O(c \log c)$.

From [Lemma 3.6](#) (by setting $\delta = 1/c^4$) we get that there is a decision tree for f of depth at most $O(c^2 \log^2 c)$ with error under μ at most 0.49. \square

3.2 Proof of [Lemma 3.6](#)

In this section, we show that if a product distribution μ is feasible, then the functions admit a decision tree of low complexity. This decision tree is obtained from the feasible solution of the LP as follows. We first show that feasibility implies the existence of a biased subcube of small support (see [Claim 3.7](#)). After querying the support of this subcube, one is left with several subproblems. One of the subproblems corresponds to the subcube itself, in which case we answer according to its bias. For each of the other subproblems, we observe that the induced distribution μ admits a feasible solution consisting of rectangles with a strictly smaller support size. This is proved by showing that the contribution of rectangles whose supports are disjoint to the original subcube is negligible (see [Claim 3.8](#)). This step crucially uses the product nature of the distribution μ .

For a set $A \subseteq \{0, 1\}^n$, let $\mu_0(A) = \mu(A \cap f^{-1}(0))$ and $\mu_1(A) = \mu(A \cap f^{-1}(1))$; let $\mu_0 = \mu_0(\{0, 1\}^n)$ and $\mu_1 = \mu_1(\{0, 1\}^n)$.

Claim 3.7. *Suppose $\mu : \{0, 1\}^n \rightarrow [0, 1]$ is a product probability distribution satisfying [\(3.1\)](#), [\(3.2\)](#) and [\(3.3\)](#). Further, suppose $\delta > 0$ is such that*

$$(1 - \alpha_0)\mu_0 - \left(\frac{\beta_0}{\delta}\right)\mu_1 > 0. \quad (3.8)$$

Then, there is subcube A with support of size at most a , such that $\mu_1(A) \leq \delta\mu_0(1)$.

Proof. We say A is biased if $\mu_1(A) \leq \delta\mu_0(A)$. From [\(3.2\)](#), we have

$$\beta_0\mu_1 \geq \sum_A \mu_1(A)u_A \quad (3.9)$$

$$\geq \sum_{A \text{ not biased}} \mu_1(A)u_A \quad (3.10)$$

$$\geq \delta \sum_{A \text{ not biased}} \mu_0(A)u_A. \quad (3.11)$$

Combining this with (3.1), we obtain

$$(1 - \alpha_0)\mu_0 - \left(\frac{\beta_0}{\delta}\right)\mu_1 \leq \sum_A u_A \mu_0(A) - \sum_{A \text{ unbiased}} u_A \mu_0(A) \quad (3.12)$$

$$\leq \sum_{A \text{ biased}} u_A \mu_0(A). \quad (3.13)$$

Since the left hand side is positive, the sum on the right cannot be empty. The claim follows from this. \square

Claim 3.8. *Let $\delta > 0$. Fix a product distribution μ , and let A be a subcube such that $\mu_1(A) \leq \delta\mu_0(A)$. Suppose $(w_R : R \text{ a subcube})$ satisfies (3.5) and (3.6). Let $\mathcal{B} = \{B : \text{support}(B) \cap \text{support}(A) = \emptyset\}$. Then,*

$$\sum_{B \in \mathcal{B}} \mu_1(B)w_B \leq \beta_1 + \delta.$$

Proof. We will use the product nature of the distribution in the following form: if A and B are subcubes with disjoint supports, then $\mu(A)\mu(B) = \mu(A \cap B)$. Thus, we have

$$\sum_{B \in \mathcal{B}} \mu_1(B)w_B \leq \sum_{B \in \mathcal{B}} \mu(B)w_B \quad (3.14)$$

$$\leq \frac{1}{\mu(A)} \sum_{B \in \mathcal{B}} \mu(A \cap B)w_B \quad (3.15)$$

$$\leq \frac{1}{\mu(A)} \sum_{B \in \mathcal{B}} \mu_0(A \cap B)w_B + \frac{1}{\mu(A)} \sum_{B \in \mathcal{B}} \mu_1(A \cap B)w_B. \quad (3.16)$$

Let us bound the two terms on the right separately. For the first term, by our assumption (3.6), we have

$$\sum_{B \in \mathcal{B}} \mu_0(A \cap B)w_B \leq \beta_1 \mu_0(A) \leq \beta_1 \mu(A).$$

For second term, by assumption (3.5), we have

$$\sum_{B \in \mathcal{B}} \mu_1(A \cap B)w_B \leq \mu_1(A) \leq \delta\mu_0(A) \leq \delta\mu(A).$$

By using these bounds in (3.16), we establish our claim. \square

Proof of Lemma 3.6. We assume that $\mu_0, \mu_1 \geq \frac{1}{4}$, for otherwise, we can reliably guess the answer without making any query. Similarly, we assume that (3.8) holds, for otherwise, we may answer 1 without making any query, and yet err with probability at most

$$\mu_0 \leq \frac{\beta_0 \mu_1}{(1 - \alpha_0)\delta}.$$

We now proceed by induction on b .

Base case ($b = 0$): The only subcube R that may appear in the inequalities (3.4)–(3.6) is the one with empty support (that is, R contains all inputs). Since $\mu_1 \geq \frac{1}{4}$, we conclude from (3.4) and (3.6) that $1 - \alpha_1 \leq w_R \leq \beta$ or $\alpha_1 + \beta_1 \geq 1$; so the claim holds trivially.

Induction step ($b \geq 1$): Using (3.1) and (3.2) and Claim 3.7, we conclude that there is a rectangle A_0 such that $\mu_1(A_0) \leq \delta\mu_0(A_0)$. We first query the bits in the support of A_0 . For each result $\sigma \in \{0, 1\}^a$, we are left with a subcube of inputs to investigate.

By Claim 3.8, we have, as B ranges over subcubes whose supports are disjoint from A_0 's, that

$$\sum_B \mu_1(B)w_B \leq \beta_1 + \delta.$$

It follows from (3.4) that (now summing over all R whose supports intersect A_0 's)

$$\begin{aligned} \sum_R \mu_1(R)w_R &\geq (1 - \alpha_1)\mu_1 - \beta_1 - \delta \\ &\geq (1 - (\alpha_1 + 4\beta_1 + 4\delta))\mu_1. \end{aligned} \tag{3.17}$$

For each outcome σ for the bits queried, let μ^σ be the resulting conditional distribution on inputs, where variables in support of A_0 are fixed at σ in μ^σ . We will now construct an LP-solution satisfying (3.1)–(3.4) for this derived problem. The components u_R will be retained without any change. For w_R , set $w_R = 0$ for R whose support is disjoint from A_0 's, and define α_1^σ by

$$\sum_R \mu_1^\sigma(R)w_R := (1 - \alpha_1^\sigma)\mu_1^\sigma.$$

Then, by (3.17), μ^σ is an $(\alpha_0, \beta_0, \alpha_1^\sigma, \beta_1, a, b - 1)$ -product distribution for f (recall our convention that we do not include the index of a fixed bit in the support of our subcubes). Furthermore, by (3.17)

$$\mathbb{E}[\alpha_1^\sigma] \leq \alpha_1 + 4\beta_1 + 4\delta. \tag{3.18}$$

By induction, μ^σ admits a decision tree of depth at most $a(b - 1)$ that errs with probability at most

$$\epsilon^\sigma \leq \frac{1}{4} + \alpha_1^\sigma + \beta_1 + 4(b - 1)(\beta_1 + \delta) + \frac{\beta_0}{(1 - \alpha_0)\delta},$$

when inputs are drawn according to μ^σ . It follows from (3.18), the overall error is

$$\begin{aligned} \mathbb{E}_\sigma[\epsilon^\sigma] &\leq \mathbb{E}_\sigma \left[\frac{1}{4} + \alpha_1^\sigma + \beta_1 + 4(b - 1)(\beta_1 + \delta) + \frac{\beta_0}{(1 - \alpha_0)\delta} \right] \\ &\leq \frac{1}{4} + \alpha_1 + \beta_1 + 4b(\beta_1 + \delta) + \frac{\beta_0}{(1 - \alpha_0)\delta}. \end{aligned} \quad \square$$

Acknowledgements

We thank Mika Göös for pointing out an error in an earlier version of the paper and for referring us to the literature.

References

- [1] Ziv Bar-Yossef, T. S. Jayram, Ravi Kumar, and D. Sivakumar. An information statistics approach to data stream and communication complexity. *J. Comput. Syst. Sci.*, 68(4):702–732, June 2004. (Preliminary version in 43rd FOCS, 2002). doi:10.1016/j.jcss.2003.11.006. 18

- [2] Boaz Barak, Mark Braverman, Xi Chen, and Anup Rao. How to compress interactive communication. *SIAM J. Comput.*, 42(3):1327–1363, 2013. (Preliminary Version in *42nd STOC*, 2010). [doi:10.1137/100811969](#). 1, 19
- [3] Mark Braverman. Interactive information complexity. *SIAM J. Comput.*, 44(6):1698–1739, 2015. (Preliminary version in *44th STOC*, 2012). [eccc:TR11-123](#), [doi:10.1137/130938517](#). 19
- [4] Amit Chakrabarti, Yaoyun Shi, Anthony Wirth, and Andrew Chi-Chih Yao. Informational complexity and the direct sum problem for simultaneous message complexity. In *Proc. 42nd IEEE Symp. on Foundations of Comp. Science (FOCS)*, pages 270–278, 2001. [doi:10.1109/SFCS.2001.959901](#). 18
- [5] Lila Fontes, Rahul Jain, Iordanis Kerenidis, Sophie Laplante, Mathieu Laurière, and Jérémie Roland. Relative discrepancy does not separate information and communication complexity. In Magnús M. Halldórsson, Kazuo Iwama, Naoki Kobayashi, and Bettina Speckmann, editors, *Proc. 42nd International Colloq. of Automata, Languages and Programming (ICALP), Part I*, volume 9134 of *LNCS*, pages 506–516. Springer, 2015. [eccc:TR15-028](#), [doi:10.1007/978-3-662-47672-7_41](#). 1
- [6] Anat Ganor, Gillat Kol, and Ran Raz. Exponential separation of information and communication. In *Proc. 55th IEEE Symp. on Foundations of Comp. Science (FOCS)*, pages 176–185, 2014. [eccc:TR14-049](#), [doi:10.1109/FOCS.2014.27](#). 1
- [7] Anat Ganor, Gillat Kol, and Ran Raz. Exponential separation of information and communication for boolean functions. In *Proc. 47th ACM Symp. on Theory of Computing (STOC)*, pages 557–566, 2015. [eccc:TR14-113](#), [doi:10.1145/2746539.2746572](#). 1
- [8] Anat Ganor, Gillat Kol, and Ran Raz. Exponential separation of communication and external information. In *Proc. 48th ACM Symp. on Theory of Computing (STOC)*, 2016. (to appear). [eccc:TR15-088](#). 1
- [9] Mika Göös, T. S. Jayram, Toniann Pitassi, and Thomas Watson. Randomized communication vs. partition number. Technical Report TR15-169, *Elect. Colloq. on Comput. Complexity (ECCC)*, 2015. URL: <http://eccc.hpi-web.de/report/2015/169>. 1
- [10] Mika Göös, Shachar Lovett, Raghu Meka, Thomas Watson, and David Zuckerman. Rectangles are nonnegative juntas. In *Proc. 47th ACM Symp. on Theory of Computing (STOC)*, pages 257–266, 2015. [eccc:TR14-147](#), [doi:10.1145/2746539.2746596](#). 4
- [11] Rahul Jain and Hartmut Klauck. The partition bound for classical communication complexity and query complexity. In *Proc. 25th IEEE Conf. on Computational Complexity*, pages 247–258, 2010. [arXiv:0910.4266](#), [doi:10.1109/CCC.2010.31](#). 1, 3, 12, 17, 19, 20
- [12] Rahul Jain and Penghui Yao. A strong direct product theorem in terms of the smooth rectangle bound. 2012. [arXiv:1209.0263](#). 3
- [13] Iordanis Kerenidis, Sophie Laplante, Virginie Lerays, Jérémie Roland, and David Xiao. Lower bounds on information complexity via zero-communication protocols and applications. *SIAM J. Comput.*, 44(5):1550–1572, 2015. (Preliminary version in *53rd FOCS*, 2012). [arXiv:1204.1505](#), [doi:10.1137/130928273](#). 20
- [14] Gillat Kol. Interactive compression for product distributions. In *Proc. 48th ACM Symp. on Theory of Computing (STOC)*, 2016. (to appear). [eccc:TR15-168](#). 1
- [15] Eyal Kushilevitz and Noam Nisan. *Communication Complexity*. Cambridge University Press, 1997. [doi:10.1017/CB09780511574948](#). 2, 4, 8
- [16] Noam Nisan and Avi Wigderson. Rounds in communication complexity revisited. *SIAM J. Comput.*, 22(1):211–219, February 1993. (Preliminary version in *23rd STOC*, 1991). [doi:10.1137/0222016](#). 18
- [17] Noam Nisan and Avi Wigderson. On rank vs. communication complexity. *Combinatorica*, 15(4):557–565, 1995. (Preliminary version in *35th FOCS*, 1994). [doi:10.1007/BF01192527](#). 2, 4

- [18] Stephen Ponzio, Jaikumar Radhakrishnan, and Srinivasan Venkatesh. The communication complexity of pointer chasing. *J. Comput. Syst. Sci.*, 62(2):323–355, 2001. (Preliminary version in *31st STOC*, 1999). doi:10.1006/jcss.2000.1731. 18
- [19] Clifford D. Smyth. Reimer’s inequality and Tardos’ conjecture. In *Proc. 34th ACM Symp. on Theory of Computing (STOC)*, pages 218–221, 2002. doi:10.1145/509907.509942. 2
- [20] Andrew Chi-Chih Yao. Some complexity questions related to distributive computing (preliminary report). In *Proc. 11th ACM Symp. on Theory of Computing (STOC)*, pages 209–213, 1979. doi:10.1145/800135.804414. 2
- [21] Andrew Chi-Chih Yao. Lower bounds by probabilistic arguments (extended abstract). In *Proc. 24th IEEE Symp. on Foundations of Comp. Science (FOCS)*, pages 420–428, 1983. doi:10.1109/SFCS.1983.30. 2

A Query Complexity LP bounds

Definition A.1 (Query partition bound [11]). *The query partition bound of $f : \{0, 1\}^n \rightarrow \{0, 1\}$, denoted $\text{qprt}_\varepsilon(f)$, is the optimal value of the following linear program. In the following A is a subcube of $\{0, 1\}^n$.*

<u>Primal</u>	<u>Dual</u>
$\min \sum_z \sum_A w_{z,A} \cdot 2^{ A }$ <p style="text-align: right; margin-right: 50px;">(A.1)</p>	$\max \quad (1 - \varepsilon) \sum_x \mu_x + \sum_x \varphi_x$
<p>s.t.</p> $\forall x \in \{0, 1\}^n : \sum_{A: x \in A} w_{f(x), A} \geq 1 - \varepsilon$ <p style="text-align: right; margin-right: 50px;">(A.2)</p>	<p>s.t.</p> $\forall (z, A) : \sum_{x \in A \cap f^{-1}(z)} \mu_x + \sum_{x \in A} \varphi_x \leq 2^{ A }$
$\forall x \in \{0, 1\}^n : \sum_{A: x \in A} \sum_z w_{z,A} = 1$ <p style="text-align: right; margin-right: 50px;">(A.3)</p>	$\forall x : \mu_x \geq 0, \varphi_x \in \mathbb{R}$
$\forall (z, A) : w_{z,A} \geq 0$ <p style="text-align: right; margin-right: 50px;">(A.4)</p>	

The error parameter for partition bound can be reduced as in the following claim.

Claim A.2. *Let $\varepsilon > \delta > 0$. Then $\log \text{qprt}_\delta(f) \leq O\left(\frac{1}{(0.5-\varepsilon)^2} \log \frac{1}{\delta}\right) \log \text{qprt}_\varepsilon(f)$.*

Proof. Let $\{w_{z,A}\}$ be an optimal solution for the primal of $\text{qprt}_\varepsilon(f)$. Using this solution, we will obtain a solution $\{v_{z,A}\}$ showing that $\text{qprt}_\delta(f)$ is small. Let t be an odd positive integer. For $z \in \{0, 1\}$, let $G_z = \{(z_1, \dots, z_t) : \sum_i z_i \geq t/2\}$ and for a subcube A , let $G_A = \{(A_1, \dots, A_t) : \bigcap_i A_i = A\}$. Now consider the following assignment to the variables:

$$v_{z,A} := \sum_{(z_1, \dots, z_t) \in G_z, (A_1, \dots, A_t) \in G_A} \prod_{i=1}^t w_{z_i, A_i}.$$

We will choose t such that this assignment constitutes a valid solution to the above linear program with δ instead of ε .

Fix x . Then,

$$\begin{aligned}
\sum_{A \ni x} v_{f(x), A} &= \sum_{A \ni x} \sum_{(z_1, \dots, z_t) \in G_{f(x)}, (A_1, \dots, A_t) \in G_A} \prod_{i=1}^t w_{z_i, A_i} \\
&= \sum_{(z_1, \dots, z_t) \in G_{f(x)}, (A_1 \ni x, \dots, A_t \ni x)} \prod_{i=1}^t w_{z_i, A_i} \\
&\geq 1 - \exp(-2(0.5 - \varepsilon)^2 t). \quad (\text{using Chernoff bounds})
\end{aligned}$$

By choosing an appropriate $t = O\left(\frac{1}{(0.5 - \varepsilon)^2} \ln \frac{1}{\delta}\right)$, we ensure that constraint (A.2) is satisfied, with δ instead of ε . Furthermore,

$$\begin{aligned}
\sum_z \sum_{A \ni x} v_{z, A} &= \sum_z \sum_{A \ni x} \sum_{(z_1, \dots, z_t) \in G_z, (A_1, \dots, A_t) \in G_A} (\prod_{i=1}^t w_{z_i, A_i}) \\
&= \sum_{(z_1, \dots, z_t), (A_1 \ni x, \dots, A_t \ni x)} (\prod_{i=1}^t w_{z_i, A_i}) \\
&= \prod_{i=1}^t \sum_{z_i} \sum_{A_i \ni x} w_{z_i, A_i} \\
&= 1.
\end{aligned}$$

Thus, the constraint (A.3) is satisfied. Constraint (A.4) is clearly satisfied.

It remains to estimate the new objective value.

$$\begin{aligned}
\sum_z \sum_A v_{z, A} 2^{|A|} &\leq \sum_z \sum_A \sum_{(z_1, \dots, z_t) \in G_z, (A_1, \dots, A_t) \in G_A} \left(\prod_{i=1}^t w_{z_i, A_i} 2^{|A_i|} \right) \\
&= \sum_{(z_1, \dots, z_t), (A_1, \dots, A_t)} \left(\prod_{i=1}^t w_{z_i, A_i} 2^{|A_i|} \right) \\
&= \prod_{i=1}^t \left(\sum_{z_i} \sum_{A_i} w_{z_i, A_i} 2^{|A_i|} \right).
\end{aligned}$$

Thus, the right hand side is at most $\text{qprt}_\varepsilon(f)^t$. Our claim follows by taking logs. \square

B Communication Complexity LP bounds

Information Complexity The notion of information complexity was formalized by Chakrabarti, Shi, Wirth and Yao [4] in the direct sum context. Similar information theoretic arguments have been used in earlier works as well without explicitly defining information complexity (see Nisan and Wigderson [16] and Ponzio, Radhakrishnan and Venkatesh [18] for applications to the pointer-chasing problem). Chakrabarti *et al.* [4] defined and used, what in today's language is called, "external information cost". Bar-Yossef, Jayram, Kumar and Sivakumar [1] used the notion, what in today's language is called the "internal information cost", in their proof of the disjointness lower bound.

The formal definition and the terminology “internal information cost” and “external information cost” was introduced by Barak *et al.* [2].

Definition B.1 (Information Complexity). Let Π be any randomized protocol between two parties Alice and Bob with inputs $\mathcal{X} \times \mathcal{Y}$ that are distributed according to a distribution μ . The (internal) information cost of the protocol Π , denoted $\text{IC}^\mu(\Pi)$ is defined as

$$\text{IC}^\mu(\Pi) = I[X : T_\Pi | Y] + I[Y : T_\Pi | X],$$

where (X, Y) is the random variable denoting the pair of inputs and T_Π is the random variable denoting the transcript of the protocol (including public randomness).

For a function $f : \mathcal{X} \times \mathcal{Y} \rightarrow \mathcal{Z}$, $\varepsilon \in (0, 1)$ and a distribution μ on $\mathcal{X} \times \mathcal{Y}$, the ε -information complexity of f under μ , denoted $\text{IC}_\varepsilon^\mu(f)$, is defined to be $\min_{\Pi} \text{IC}^\mu(\Pi)$, where the minimum is taken over all protocols Π that compute f with error at most ε under the distribution μ . The ε -information complexity of f , denoted $\text{IC}_\varepsilon(f)$, is defined to be $\max_\mu \text{IC}_\varepsilon^\mu(f)$, where the maximum is over all distributions μ .

An alternate notion of information complexity was considered by Braverman [3] by changing the order of quantifiers.

Definition B.2 (Prior-free information complexity). For a function $f : \mathcal{X} \times \mathcal{Y} \rightarrow \mathcal{Z}$ and $\varepsilon \in (0, 1)$, the prior free information complexity, denoted $\text{IC}_\varepsilon^p(f)$, is defined to be $\min_{\Pi} \max_\mu \text{IC}^\mu(\Pi)$ where the minimum is taken over all randomized protocols Π that compute the function f correctly on all inputs with error at most ε and the maximum is over all distributions μ .

Braverman [3] showed that these two notions of information complexity are within constant factors of each other. More precisely,

$$\text{IC}_\varepsilon(f) \leq \text{IC}_\varepsilon^p(f) \leq 2 \cdot \text{IC}_{\varepsilon/2}(f).$$

This alternate characterization of information complexity immediately yields the following error reduction claim.

Claim B.3. Let $\delta > 0$. Then $\text{IC}_\delta(f) \leq O(\log \frac{1}{\delta}) \cdot \text{IC}_{1/8}(f)$.

The partition bound: The partition bound was introduced by Jain and Klauck [11] as a linear programming bound to lower bound the public-coin randomized communication complexity.

Definition B.4 (Partition bound [11]). For a function $f : \mathcal{X} \times \mathcal{Y} \rightarrow \mathcal{Z}$ and $\varepsilon \in (0, 1)$, the ε -partition bound of f , denoted $\text{qprt}_\varepsilon(f)$, is defined to be the optimal value of the following linear program. Below R represents a rectangle in $\mathcal{X} \times \mathcal{Y}$ and $(x, y, z) \in \mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$.

<u>Primal</u>	<u>Dual</u>
$\begin{aligned} \min \quad & \sum_z \sum_R w_{z,R} \\ & \sum_{R: (x,y) \in R} w_{f(x,y),R} \geq 1 - \varepsilon, \quad \forall (x,y) \\ & \sum_{R: (x,y) \in R} \sum_z w_{z,R} = 1, \quad \forall (x,y) \\ & w_{z,R} \geq 0, \quad \forall (z,R) . \end{aligned}$	$\begin{aligned} \max \quad & (1 - \varepsilon) \sum_{(x,y)} \mu_{x,y} + \sum_{(x,y)} \varphi_{x,y} \\ & \sum_{(x,y) \in R \cap f^{-1}(z)} \mu_{x,y} + \sum_{(x,y) \in R} \varphi_{x,y} \leq 1, \quad \forall (z,R) \\ & \mu_{x,y} \geq 0, \quad \forall (x,y) \\ & \varphi_{x,y} \in \mathbb{R}, \quad \forall (x,y) . \end{aligned}$

Jain and Klauck [11] show that $\text{CC}_\varepsilon^{\text{pub}}(f) \geq \log(\text{prt}_\varepsilon(f))$.

The error in the partition bound in the communication complexity setting can be reduced in a fashion similar to query setting (ie., Claim A.2).

Claim B.5. *Let $\delta > 0$. Then $\log \text{prt}_\delta(f) \leq O(\log \frac{1}{\delta}) \cdot \log \text{prt}_{1/8}(f)$.*

Relaxed partition bound: Kerenidis *et al.* [13] defined a relaxation of the partition bound by relaxing the equality constraint in the primal.

Definition B.6 (Relaxed partition bound [13]). *For a function $f : \mathcal{X} \times \mathcal{Y} \rightarrow \mathcal{Z}$ and $\varepsilon \in (0, 1)$, the ε -relaxed partition bound of f , denoted $\text{rprt}_\varepsilon(f)$, is given by the optimal value of the following linear program.*

<u>Primal</u>	<u>Dual</u>
$\begin{aligned} \min \quad & \sum_z \sum_R w_{z,R} \\ & \sum_{R: (x,y) \in R} w_{f(x,y),R} \geq 1 - \varepsilon, \quad \forall (x,y) \\ & \sum_{R: (x,y) \in R} \sum_z w_{z,R} \leq 1, \quad \forall (x,y) \\ & w_{z,R} \geq 0, \quad \forall (z,R) . \end{aligned}$	$\begin{aligned} \max \quad & (1 - \varepsilon) \sum_{(x,y)} \mu_{x,y} - \sum_{(x,y)} \varphi_{x,y} \\ & \sum_{(x,y) \in R \cap f^{-1}(z)} \mu_{x,y} - \sum_{(x,y) \in R} \varphi_{x,y} \leq 1, \quad \forall (z,R) \\ & \mu_{x,y} \geq 0, \quad \forall (x,y) \\ & \varphi_{x,y} \geq 0, \quad \forall (x,y) . \end{aligned}$

Clearly, $\text{CC}_\varepsilon^{\text{pub}}(f) \geq \log(\text{prt}_\varepsilon(f)) \geq \log(\text{rprt}_\varepsilon(f)) \geq \log \text{srec}_{\varepsilon,\varepsilon}(f)$. Kerenidis *et al.* [13] showed that relaxed partition bound is upper bounded by information complexity. Combining these facts with the error reduction claims for partition bound and information complexity (Claims B.3 and B.5) yields Proposition 2.5.